SEQUENCE LISTING PART OF THE DESCRIPTION

SEQ. ID. NO. 1 - Wild type gagpol sequence for strain HXB2 (accession no. K03455)

```
ATGGGTGCGA GAGCGTCAGT ATTAAGCGGG GGAGAATTAG ATCGATGGGA AAAAATTCGG 60
TTAAGGCCAG GGGGAAAGAA AAAATATAAA TTAAAACATA TAGTATGGGC AAGCAGGGAG 120
CTAGAACGAT TCGCAGTTAA TCCTGGCCTG TTAGAAACAT CAGAAGGCTG TAGACAAATA 180
CTGGGACAGC TACAACCATC CCTTCAGACA GGATCAGAAG AACTTAGATC ATTATATAAT 240
ACAGTAGCAA CCCTCTATTG TGTGCATCAA AGGATAGAGA TAAAAGACAC CAAGGAAGCT 300
TTAGACAAGA TAGAGGAAGA GCAAAACAAA AGTAAGAAAA AGCACAGCA AGCAGCAGCT 360
GACACAGGAC ACAGCAATCA GGTCAGCCAA AATTACCCTA TAGTGCAGAA CATCCAGGGG 420
CAAATGGTAC ATCAGGCCAT ATCACCTAGA ACTTTAAATG CATGGGTAAA AGTAGTAGAA 480
GAGAAGGCTT TCAGCCCAGA AGTGATACCC ATGTTTTCAG CATTATCAGA AGGAGCCACC 540
CCACAAGATT TAAACACCAT GCTAAACACA GTGGGGGGAC ATCAAGCAGC CATGCAAATG 600
TTAAAAGAGA CCATCAATGA GGAAGCTGCA GAATGGGATA GAGTGCATCC AGTGCATGCA 660
GGGCCTATTG CACCAGGCCA GATGAGAGAA CCAAGGGGAA GTGACATAGC AGGAACTACT 720
AGTACCCTTC AGGAACAAAT AGGATGGATG ACAAATAATC CACCTATCCC AGTAGGAGAA 780
ATTTATAAAA GATGGATAAT CCTGGGATTA AATAAAATAG TAAGAATGTA TAGCCCTACC 840
AGCATTCTGG ACATAAGACA AGGACCAAAG GAACCCTTTA GAGACTATGT AGACCGGTTC 900
TATAAAACTC TAAGAGCCGA GCAAGCTTCA CAGGAGGTAA AAAATTGGAT GACAGAAACC 960
TTGTTGGTCC AAAATGCGAA CCCAGATTGT AAGACTATTT TAAAAGCATT GGGACCAGCG 1020
GCTACACTAG AAGAAATGAT GACAGCATGT CAGGGAGTAG GAGGACCCGG CCATAAGGCA 1080
AGAGTTTTGG CTGAAGCAAT GAGCCAAGTA ACAAATTCAG CTACCATAAT GATGCAGAGA 1140
GGCAATTTTA GGAACCAAAG AAAGATTGTT AAGTGTTTCA ATTGTGGCAA AGAAGGGCAC 1200
ACAGCCAGAA ATTGCAGGGC CCCTAGGAAA AAGGGCTGTT GGAAATGTGG AAAGGAAGGA 1260
CACCAAATGA AAGATTGTAC TGAGAGACAG GCTAATTTTT TAGGGAAGAT CTGGCCTTCC 1320
TACAAGGGAA GGCCAGGGAA TTTTCTTCAG AGCAGACCAG AGCCAACAGC CCCACCAGAA 1380
GAGAGCTTCA GGTCTGGGGT AGAGACAACA ACTCCCCCTC AGAAGCAGGA GCCGATAGAC 1440
AAGGAACTGT ATCCTTTAAC TTCCCTCAGG TCACTCTTTG GCAACGACCC CTCGTCACAA 1500
TAAAGATAGG GGGGCAACTA AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT 1560
TAGAAGAAAT GAGTTTGCCA GGAAGATGGA AACCAAAAAT GATAGGGGGA ATTGGAGGTT 1620
TTATCAAAGT AAGACAGTAT GATCAGATAC TCATAGAAAT CTGTGGACAT AAAGCTATAG 1680
GTACAGTATT AGTAGGACCT ACACCTGTCA ACATAATTGG AAGAAATCTG TTGACTCAGA 1740
TTGGTTGCAC TTTAAATTTT CCCATTAGCC CTATTGAGAC TGTACCAGTA AAATTAAAGC 1800
CAGGAATGGA TGGCCCAAAA GTTAAACAAT GGCCATTGAC AGAAGAAAAA ATAAAAGCAT 1860
TAGTAGAAAT TTGTACAGAG ATGGAAAAGG AAGGGAAAAT TTCAAAAATT GGGCCTGAAA 1920
ATCCATACAA TACTCCAGTA TTTGCCATAA AGAAAAAAGA CAGTACTAAA TGGAGAAAAT 1980
TAGTAGATTT CAGAGAACTT AATAAGAGAA CTCAAGACTT CTGGGAAGTT CAATTAGGAA 2040
TACCACATCC CGCAGGGTTA AAAAAGAAAA AATCAGTAAC AGTACTGGAT GTGGGTGATG 2100
CATATTTTTC AGTTCCCTTA GATGAAGACT TCAGGAAGTA TACTGCATTT ACCATACCTA 2160
GTATAAACAA TGAGACACCA GGGATTAGAT ATCAGTACAA TGTGCTTCCA CAGGGATGGA 2220
AAGGATCACC AGCAATATTC CAAAGTAGCA TGACAAAAAT CTTAGAGCCT TTTAGAAAAC 2280
AAAATCCAGA CATAGTTATC TATCAATACA TGGATGATTT GTATGTAGGA TCTGACTTAG 2340
AAATAGGGCA GCATAGAACA AAAATAGAGG AGCTGAGACA ACATCTGTTG AGGTGGGGAC 2400
TTACCACACC AGACAAAAAA CATCAGAAAG AACCTCCATT CCTTTGGATG GGTTATGAAC 2460
TCCATCCTGA TAAATGGACA GTACAGCCTA TAGTGCTGCC AGAAAAAGAC AGCTGGACTG 2520
TCAATGACAT ACAGAAGTTA GTGGGGAAAT TGAATTGGGC AAGTCAGATT TACCCAGGGA 2580
TTAAAGTAAG GCAATTATGT AAACTCCTTA GAGGAACCAA AGCACTAACA GAAGTAATAC 2640
CACTAACAGA AGAAGCAGAG CTAGAACTGG CAGAAAACAG AGAGATTCTA AAAGAACCAG 2700
TACATGGAGT GTATTATGAC CCATCAAAAG ACTTAATAGC AGAAATACAG AAGCAGGGGC 2760
AAGGCCAATG GACATATCAA ATTTATCAAG AGCCATTTAA AAATCTGAAA ACAGGAAAAT 2820
ATGCAAGAAT GAGGGGTGCC CACACTAATG ATGTAAAACA ATTAACAGAG GCAGTGCAAA 2880
AAATAACCAC AGAAAGCATA GTAATATGGG GAAAGACTCC TAAATTTAAA CTGCCCATAC 2940
AAAAGGAAAC ATGGGAAACA TGGTGGACAG AGTATTGGCA AGCCACCTGG ATTCCTGAGT 3000
GGGAGTTTGT TAATACCCCT CCCTTAGTGA AATTATGGTA CCAGTTAGAG AAAGAACCCA 3060
TAGTAGGAGC AGAAACCTTC TATGTAGATG GGGCAGCTAA CAGGGAGACT AAATTAGGAA 3120
AAGCAGGATA TGTTACTAAT AGAGGAAGAC AAAAAGTTGT CACCCTAACT GACACAACAA 3180
ATCAGAAGAC TGAGTTACAA GCAATTTATC TAGCTTTGCA GGATTCGGGA TTAGAAGTAA 3240
ACATAGTAAC AGACTCACAA TATGCATTAG GAATCATTCA AGCACAACCA GATCAAAGTG 3300
AATCAGAGTT AGTCAATCAA ATAATAGAGC AGTTAATAAA AAAGGAAAAG GTCTATCTGG 3360
CATGGGTACC AGCACACAAA GGAATTGGAG GAAATGAACA AGTAGATAAA TTAGTCAGTG 3420
CTGGAATCAG GAAAGTACTA TTTTTAGATG GAATAGATAA GGCCCAAGAT GAACATGAGA 3480
AATATCACAG TAATTGGAGA GCAATGGCTA GTGATTTTAA CCTGCCACCT GTAGTAGCAA 3540
```

```
AAGAAATAGT AGCCAGCTGT GATAAATGTC AGCTAAAAGG AGAAGCCATG CATGGACAAG 3600
TAGACTGTAG TCCAGGAATA TGGCAACTAG ATTGTACACA TTTAGAAGGA AAAGTTATCC 3660
TGGTAGCAGT TCATGTAGCC AGTGGATATA TAGAAGCAGA AGTTATTCCA GCAGAAACAG 3720
GGCAGGAAAC AGCATATTTT CTTTTAAAAT TAGCAGGAAG ATGGCCAGTA AAAACAATAC 3780
ATACTGACAA TGGCAGCAAT TTCACCGGTG CTACGGTTAG GGCCGCCTGT TGGTGGGCGG 3840
GAATCAAGCA GGAATTTGGA ATTCCCTACA ATCCCCAAAG TCAAGGAGTA GTAGAATCTA 3900
TGAATAAAGA ATTAAAGAAA ATTATAGGAC AGGTAAGAGA TCAGGCTGAA CATCTTAAGA 3960
CAGCAGTACA AATGGCAGTA TTCATCCACA ATTTTAAAAG AAAAGGGGGG ATTGGGGGGT 4020
ACAGTGCAGG GGAAAGAATA GTAGACATAA TAGCAACAGA CATACAAACT AAAGAATTAC 4080
AAAAACAAAT TACAAAAATT CAAAATTTTC GGGTTTATTA CAGGGACAGC AGAAATTCAC 4140
TTTGGAAAGG ACCAGCAAAG CTCCTCTGGA AAGGTGAAGG GGCAGTAGTA ATACAAGATA 4200
ATAGTGACAT AAAAGTAGTG CCAAGAAGAA AAGCAAAGAT CATTAGGGAT TATGGAAAAC 4260
AGATGGCAGG TGATGATTGT GTGGCAAGTA GACAGGATGA GGATTAG              4307
```

SEQ I.D. NO. 2 - gagpol-SYNgp - codon optimised gagpol sequence

```
ATGGGCGCCC GCGCCAGCGT GCTGTCGGGC GGCGAGCTGG ACCGCTGGGA GAAGATCCGC 60
CTGCGCCCCG GCGGCAAAAA GAAGTACAAG CTGAAGCACA TCGTGTGGGC CAGCCGCGAA 120
CTGGAGCGCT TCGCCGTGAA CCCCGGGCTC CTGGAGACCA GCGAGGGGTG CCGCCAGATC 180
CTCGGCCAAC TGCAGCCCAG CCTGCAAACG GGCAGCGAGG AGCTGCGCAG CCTGTACAAC 240
ACCGTGGCCA CGCTGTACTG CGTCCACCAG CGCATCGAAA TCAAGGATAC GAAAGAGGCC 300
CTGGATAAAA TCGAAGAGGA ACAGAATAAG AGCAAAAAGA AGGCCCAACA GGCCGCCGCG 360
GACACCGGAC ACAGCAACCA GGTCAGCCAG AACTACCCCA TCGTGCAGAA CATCCAGGGG 420
CAGATGGTGC ACCAGGCCAT CTCCCCCCGC ACGCTGAACG CCTGGGTGAA GGTGGTGGAA 480
GAGAAGGCTT TTAGCCCGGA GGTGATACCC ATGTTCTCAG CCCTGTCAGA GGGAGCCACC 540
CCCCAAGATC TGAACACCAT GCTCAACACA GTGGGGGGAC ACCAGGCCGC CATGCAGATG 600
CTGAAGGAGA CCATCAATGA GGAGGCTGCC GAATGGGATC GTGTGCATCC GGTGCACGCA 660
GGGCCCATCG CACCGGGCCA GATGCGTGAG CCACGGGGCT CAGACATCGC CGGAACGACT 720
AGTACCCTTC AGGAACAGAT CGGCTGGATG ACCAACAACC CACCCATCCC GGTGGGAGAA 780
ATCTACAAAC GCTGGATCAT CCTGGGCCTG AACAAGATCG TGCGCATGTA TAGCCCTACC 840
AGCATCCTGG ACATCCGCCA AGGCCCGAAG GAACCCTTTC GCGACTACGT GGACCGGTTC 900
TACAAAACGC TCCGCGCCGA GCAGGCTAGC CAGGAGGTGA AGAACTGGAT GACCGAAACC 960
CTGCTGGTCC AGAACGCGAA CCCCGGACTGC AAGACGATCC TGAAGGCCCT GGGCCCAGCG 1020
GCTACCCTAG AGGAAATGAT GACCGCCTGT CAGGGAGTGG CGGACCCGG CCACAAGGCA 1080
CGCGTCCTGG CTGAGGCCAT GAGCCAGGTG ACCAACTCCG CTACCATCAT GATGCAGCGC 1140
GGCAACTTTC GGAACCAACG CAAGATCGTC AAGTGCTTCA ACTGTGGCAA AGAAGGGCAC 1200
ACAGCCCGCA ACTGCAGGGC CCCTAGGAAA AAGGGCTGCT GGAAATGCGG CAAGGAAGGC 1260
CACCAGATGA AAGACTGTAC TGAGAGACAG GCTAATTTTT TAGGGAAGAT CTGGCCTTCC 1320
TACAAGGGAA GGCCAGGGAA TTTTCTTCAG AGCAGACCAG AGCCAACAGC CCCACCAGAA 1380
GAGAGCTTCA GGTCTGGGGT AGAGACAACA ACTCCCCCTC AGAAGCAGGA GCCGATAGAC 1440
AAGGAACTGT ATCCTTTAAC TTCCCTCAGA TCACTCTTTG GCAACGACCC CTCGTCACAA 1500
TAAAGATAGG GGGGCAGCTC AAGGAGGCTC TCCTGGACAC CGGAGCAGAC GACACCGTGC 1560
TGGAGGAGAT GTCGTTGCCA GGCCGCTGGA AGCCGAAGAT GATCGGGGGA ATCGGCGGTT 1620
TCATCAAGGT GCGCCAGTAT GACCAGATCC TCATCGAAAT CTGCGGCCAC AAGGCTATCG 1680
GTACCGTGCT GGTGGGCCCC ACACCCGTCA ACATCATCGG ACGCAACCTG TTGACGCAGA 1740
TCGGTTGCAC GCTGAACTTC CCCATTAGCC CTATCGAGAC GGTACCGGTG AAGCTGAAGC 1800
CCGGGATGGA CGGCCCGAAG GTCAAGCAAT GGCCATTGAC AGAGGAGAAG ATCAAGGCAC 1860
TGGTGGAGAT TTGCACAGAG ATGGAAAAGG AAGGGAAAAT CTCCAAGATT GGGCCTGAGA 1920
ACCCGTACAA CACGCCGGTG TTCGCAATCA AGAAGAAGGA CTCGACGAAA TGGCGCAAGC 1980
TGGTGGACTT CCGCGAGCTG AACAAGCGCA CGCAAGACTT CTGGGAGGTT CAGCTGGGCA 2040
TCCCGCACCC CGCAGGGCTG AAGAAGAAGA AATCCGTGAC CGTACTGGAT GTGGGTGATG 2100
CCTACTTCTC CGTTCCCCTG GACGAAGACT TCAGGAAGTA CACTGCCTTC ACAATCCCTT 2160
CGATCAACAA CGAGACACCG GGGATTCGAT ATCAGTACAA CGTGCTGCCC CAGGGCTGGA 2220
AAGGCTCTCC CGCAATCTTC CAGAGTAGCA TGACCAAAAT CCTGGAGCCT TTCCGCAAAC 2280
AGAACCCCGA CATCGTCATC TATCAGTACA TGGATGACTT GTACGTGGGC TCTGATCTAG 2340
AGATAGGGCA GCACCGCACC AAGATCGAGG AGCTGCGCCA GCACCTGTTG AGGTGGGGAC 2400
TGACCACACC CGACAAGAAG CACCAGAAGG AGCCTCCCTT CCTCTGGATG GGTTACGAGC 2460
TGCACCCTGA CAAATGGACC GTGCAGCCTA TCGTGCTGCC AGAGAAAGAC AGCTGGACTG 2520
TCAACGACAT ACAGAAGCTG GTGGGGAAGT TGAACTGGGC CAGTCAGATT TACCCAGGGA 2580
TTAAGGTGAG GCAGCTGTGC AAACTCCTCC GCGGAACCAA GGCACTCACA GAGGTGATCC 2640
CCCTAACCGA GGAGGCCGAG CTCGAACTGG CAGAAAACCG AGAGATCCTA AAGGAGCCCG 2700
TGCACGGCGT GTACTATGAC CCCTCCAAGG ACCTGATCGC CGAGATCCAG AAGCAGGGGC 2760
AAGGCCAGTG GACCTATCAG ATTTACCAGG AGCCCTTCAA GAACCTGAAG ACCGGCAAGT 2820
ACGCCCGGAT GAGGGGTGCC CACACTAACG ACGTCAAGCA GCTGACCGAG GCCGTGCAGA 2880
```

```
AGATCACCAC CGAAAGCATC GTGATCTGGG GAAAGACTCC TAAGTTCAAG CTGCCCATCC 2940
AGAAGGAAAC CTGGGAAACC TGGTGGACAG AGTATTGGCA GGCCACCTGG ATTCCTGAGT 3000
GGGAGTTCGT CAACACCCCT CCCCTGGTGA AGCTGTGGTA CCAGCTGGAG AAGGAGCCCA 3060
TAGTGGGCGC CGAAACCTTC TACGTGGATG GGGCCGCTAA CAGGGAGACT AAGCTGGGCA 3120
AAGCCGGATA CGTCACTAAC CGGGGCAGAC AGAAGGTTGT CACCCTCACT GACACCACCA 3180
ACCAGAAGAC TGAGCTGCAG GCCATTTACC TCGCTTTGCA GGACTCGGGC CTGGAGGTGA 3240
ACATCGTGAC AGACTCTCAG TATGCCCTGG GCATCATTCA AGCCCAGCCA GACCAGAGTG 3300
AGTCCGAGCT GGTCAATCAG ATCATCGAGC AGCTGATCAA GAAGGAAAAG GTCTATCTGG 3360
CCTGGGTACC CGCCCACAAA GGCATTGGCG GCAATGAGCA GGTCGACAAG CTGGTCTCGG 3420
CTGGCATCAG GAAGGTGCTA TTCCTGGATG GCATCGACAA GGCCCAGGAC GAGCACGAGA 3480
AATACCACAG CAACTGGCGG GCCATGGCTA GCGACTTCAA CCTGCCCCCT GTGGTGGCCA 3540
AAGAGATCGT GGCCAGCTGT GACAAGTGTC AGCTCAAGGG CGAAGCCATG CATGGCCAGG 3600
TGGACTGTAG CCCCGGCATC TGGCAACTCG ATTGCACCCA TCTGGAGGGC AAGGTTATCC 3660
TGGTAGCCGT CCATGTGGCC AGTGGCTACA TCGAGGCCGA GGTCATTCCC GCCGAAACAG 3720
GGCAGGAGAC AGCCTACTTC CTCCTGAAGC TGGCAGGCCG GTGGCCAGTG AAGACCATCC 3780
ATACTGACAA TGGCAGCAAT TTCACCAGTG CTACGGTTAA GGCCGCCTGC TGGTGGGCGG 3840
GAATCAAGCA GGAGTTCGGG ATCCCCTACA ATCCCCAGAG TCAGGGCGTC GTCGAGTCTA 3900
TGAATAAGGA GTTAAAGAAG ATTATCGGCC AGGTCAGAGA TCAGGCTGAG CATCTCAAGA 3960
CCGCGGTCCA AATGGCGGTA TTCATCCACA ATTTCAAGCG GAAGGGGGGG ATTGGGGGGT 4020
ACAGTGCGGG GGAGCGGATC GTGGACATCA TCGCGACCGA CATCCAGACT AAGGAGCTGC 4080
AAAAGCAGAT TACCAAGATT CAGAATTTCC GGGTCTACTA CAGGGACAGC AGAAATCCCC 4140
TCTGGAAAGG CCCAGCGAAG CTCCTCTGGA AGGGTGAGGG GGCAGTAGTG ATCCAGGATA 4200
ATAGCGACAT CAAGGTGGTG CCCAGAAGAA AGGCGAAGAT CATTAGGGAT TATGGCAAAC 4260
AGATGGCGGG TGATGATTGC GTGGCGAGCA GACAGGATGA GGATTAG            4307
```

SEQ. ID. NO. 3 - Envelope Gene from HIV-1 MN (Genbank accession no. M17449)

```
ATGAGAGTGA AGGGGATCAG GAGGAATTAT CAGCACTGGT GGGGATGGGG CACGATGCTC 60
CTTGGGTTAT TAATGATCTG TAGTGCTACA GAAAAATTGT GGGTCACAGT CTATTATGGG 120
GTACCTGTGT GGAAAGAAGC AACCACCACT CTATTTTGTG CATCAGATGC TAAAGCATAT 180
GATACAGAGG TACATAATGT TTGGGCCACA CAAGCCTGTG TACCCACAGA CCCCAACCCA 240
CAAGAAGTAG AATTGGTAAA TGTGACAGAA AATTTTAACA TGTGGAAAAA TAACATGGTA 300
GAACAGATGC ATGAGGATAT AATCAGTTTA TGGGATCAAA GCCTAAAGCC ATGTGTAAAA 360
TTAACCCCAC TCTGTGTTAC TTTAAATTGC ACTGATTTGA GGAATACTAC TAATACCAAT 420
AATAGTACTG CTAATAACAA TAGTAATAGC GAGGGAACAA TAAAGGGAGG AGAAATGAAA 480
AACTGCTCTT TCAATATCAC CACAAGCATA AGAGATAAGA TGCAGAAAGA ATATGCACTT 540
CTTTATAAAC TTGATATAAT ATCAATAGAT AATGATAGTA CCAGCTATAG GTTGATAAGT 600
TGTAATACCT CAGTCATTAC ACAAGCCTGT CCAAAGATAT CCTTTGAGCC AATTCCCATA 660
CACTATTGTG CCCCGGCTGG TTTTGCGATT CTAAAATGTA ACGATAAAAG GTTCAGTGGA 720
AAAGGATCAT GTAAAAATGT CAGCACAGTA CAATGTACAC ATGGAATTAG GCCAGTAGTA 780
TCAACTCAAC TGCTGTTAAA TGGCAGTCTA GCAGAAGAAG AGGTAGTAAT TAGATCTGAG 840
AATTTCACTG ATAATGCTAA AACCATCATA GTACATCTGA ATGAATCTGT ACAAATTAAT 900
TGTACAAGAC CCAACTACAA TAAAAGAAAA GGATACATA TAGGACCAGG GAGAGCATTT 960
TATACAACAA AAAATATAAT AGGAACTATA AGACAAGCAC ATTGTAACAT TAGTAGAGCA 1020
AAATGGAATG ACACTTTAAG ACAGATAGTT AGCAAATTAA AGAACAATT TAAGAATAAA 1080
ACAATAGTCT TTAATCAATC CTCAGGAGGG GACCCAGAAA TTGTAATGCA CAGTTTTAAT 1140
TGTGGAGGGG AATTTTTCTA CTGTAATACA TCACCACTGT TAATAGTAC TTGGAATGGT 1200
AATAATACTT GGAATAATAC TACAGGGTCA AATAACAATA TCACACTTCA ATGCAAAATA 1260
AAACAAATTA TAAACATGTG GCAGGAAGTA GGAAAAGCAA TGTATGCCCC TCCCATTGAA 1320
GGACAAATTA GATGTTCATC AAATATTACA GGGCTACTAT TAACAAGAGA TGGTGGTAAG 1380
GACACGGACA CGAACGACAC CGAGATCTTC AGACCTGGAG GAGGAGATAT GAGGGACAAT 1440
TGGAGAAGTG AATTATATAA ATATAAAGTA GTAACAATTG AACCATTAGG AGTAGCACCC 1500
ACCAAGGCAA AGAGAAGAGT GGTGCAGAGA GAAAAAAGAG CAGCGATAGG AGCTCTGTTC 1560
CTTGGGTTCT TAGGAGCAGC AGGAAGCACT ATGGGCGCAG CGTCAGTGAC GCTGACGGTA 1620
CAGGCCAGAC TATTATTGTC TGGTATAGTG CAACAGCAGA ACAATTTGCT GAGGGCCATT 1680
GAGGCGCAAC AGCATATGTT GCAACTCACA GTCTGGGGC TCAAGCAGCT CCAGGCAAGA 1740
GTCCTGGCTG TGGAAAGATA CCTAAAGGAT CAACAGCTCC TGGGGTTTTG GGGTTGCTCT 1800
GGAAAACTCA TTTGCACCAC TACTGTGCCT TGGAATGCTA GTTGGAGTAA TAAATCTCTG 1860
GATGATATTT GGAATAACAT GACCTGGATG CAGTGGGAAA GAGAATTGA CAATTACACA 1920
AGCTTAATAT ACTCATTACT AGAAAAATCG CAAACCCAAC AAGAAAAGAA TGAACAAGAA 1980
TTATTGGAAT TGGATAAATG GGCAAGTTTG TGGAATTGGT TTGACATAAC AAATTGGCTG 2040
TGGTATATAA AAATATTCAT AATGATAGTA GGAGGCTTGG TAGGTTTAAG AATAGTTTTT 2100
GCTGTACTTT CTATAGTGAA TAGAGTTAGG CAGGGATACT CACCATTGTC GTTGCAGACC 2160
CGCCCCCCAG TTCCGAGGGG ACCCGACAGG CCCGAAGGAA TCGAAGAAGA AGGTGGAGAG 2220
```

```
AGAGACAGAG ACACATCCGG TCGATTAGTG CATGGATTCT TAGCAATTAT CTGGGTCGAC 2280
CTGCGGAGCC TGTTCCTCTT CAGCTACCAC CACAGAGACT TACTCTTGAT TGCAGCGAGG 2340
ATTGTGGAAC TTCTGGGACG CAGGGGGTGG GAAGTCCTCA AATATTGGTG GAATCTCCTA 2400
CAGTATTGGA GTCAGGAACT AAAGAGTAGT GCTGTTAGCT TGCTTAATGC CACAGCTATA 2460
GCAGTAGCTG AGGGGACAGA TAGGGTTATA GAAGTACTGC AAAGAGCTGG TAGAGCTATT 2520
CTCCACATAC CTACAAGAAT AAGACAGGGC TTGGAAAGGG CTTTGCTATA A          2571
```

SEQ. I.D. NO. 4 - SYNgp-160mn - codon optimised env sequence

```
ATGAGGGTGA AGGGGATCCG CCGCAACTAC CAGCACTGGT GGGGCTGGGG CACGATGCTC 60
CTGGGGCTGC TGATGATCTG CAGCGCCACC GAGAAGCTGT GGGTGACCGT GTACTACGGC 120
GTGCCCGTGT GGAAGGAGGC CACCACCACC CTGTTCTGCG CCAGCGACGC CAAGGCGTAC 180
GACACCGAGG TGCACAACGT GTGGGCCACC CAGGCGTGCG TGCCCACCGA CCCCAACCCC 240
CAGGAGGTGG AGCTCGTGAA CGTGACCGAG AACTTCAACA TGTGGAAGAA CAACATGGTG 300
GAGCAGATGC ATGAGGACAT CATCAGCCTG TGGGACCAGA GCCTGAAGCC CTGCGTGAAG 360
CTGACCCCCC TGTGCGTGAC CCTGAACTGC ACCGACCTGA GGAACACCAC CAACACCAAC 420
AACAGCACCG CCAACAACAA CAGCAACAGC GAGGGCACCA TCAAGGGCGG CGAGATGAAG 480
AACTGCAGCT TCAACATCAC CACCAGCATC CGCGACAAGA TGCAGAAGGA GTACGCCCTG 540
CTGTACAAGC TGGATATCGT GAGCATCGAC AACGACAGCA CCAGCTACCG CCTGATCTCC 600
TGCAACACCA GCGTGATCAC CCAGGCCTGC CCCAAGATCA GCTTCGAGCC CATCCCCATC 660
CACTACTGCG CCCCCGCCGG CTTCGCCATC CTGAAGTGCA ACGACAAGAA GTTCAGCGGC 720
AAGGGCAGCT GCAAGAACGT GAGCACCGTG CAGTGCACCC ACGGCATCCG GCCGGTGGTG 780
AGCACCCAGC TCCTGCTGAA CGGCAGCCTG GCCGAGGAGG AGGTGGTGAT CCGCAGCGAG 840
AACTTCACCG ACAACGCCAA GACCATCATC GTGCACCTGA ATGAGAGCGT GCAGATCAAC 900
TGCACGCGTC CCAACTACAA CAAGCGCAAG CGCATCCACA TCGGCCCACGG GCGCGCCTTC 960
TACACCACCA GAAACATCAT CGGCACCATC CGCCAGGCCC ACTGCAACAT CTCTAGAGCC 1020
AAGTGGAACG ACACCCTGCG CCAGATCGTG AGCAAGCTGA AGGAGCAGTT CAAGAACAAG 1080
ACCATCGTGT TCAACCAGAG CAGCGGCGGC GACCCCGAGA TCGTGATGCA CAGCTTCAAC 1140
TGCGGCGGCG AATTCTTCTA CTGCAACACC AGCCCCCTGT TCAACAGCAC CTGGAACGGC 1200
AACAACACCT GGAACAACAC CACCGGCAGC AACAACAATA TTACCCTCCA GTGCAAGATC 1260
AAGCAGATCA TCAACATGTG GCAGGAGGTG GGCAAGGCCA TGTACGCCCC CCCCATCGAG 1320
GGCCAGATCC GGTGCAGCAG CAACATCACC GGTCTGCTGC TGACCCGCGA CGGCGGCAAG 1380
GACACCGACA CCAACGACAC CGAAATCTTC CGCCCCGGCG GCGGCGACAT GCGCGACAAC 1440
TGGAGATCTG AGCTGTACAA GTACAAGGTG GTGACGATCG AGCCCCTGGG CGTGGCCCCC 1500
ACCAAGGCCA AGCGCCGCGT GGTGCAGCGC GAGAAGCGGG CCGCCATCGG CGCCCTGTTC 1560
CTGGGCTTCC TGGGGGCGGC GGGCAGCACC ATGGGGGCCG CCAGCGTGAC CCTGACCGTG 1620
CAGGCCCGCC TGCTCCTGAG CGGCATCGTG CAGCAGCAGA ACAACCTCCT CCGCGCCATC 1680
GAGGCCCAGC AGCATATGCT CCAGCTCACC GTGTGGGGCA TCAAGCAGCT CCAGGCCCGC 1740
GTGCTGGCCG TGGAGCGCTA CCTGAAGGAC CAGCAGCTCC TGGGCTTCTG GGGCTGCTCC 1800
GGCAAGCTGA TCTGCACCAC CACGGTACCC TGGAACGCCT CCTGGAGCAA CAAGAGCCTG 1860
GACGACATCT GGAACAACAT GACCTGGATG CAGTGGGAGC GCGAGATCGA TAACTACACC 1920
AGCCTGATCT ACAGCCTGCT GGAGAAGAGC CAGCAGCAGC AGGAGAAGAA CGAGCAGGAG 1980
CTGCTGGAGC TGGACAAGTG GGCGAGCCTG TGGAACTGGT TCGACATCAC CAACTGGCTG 2040
TGGTACATCA AAATCTTCAT CATGATTGTG GGCGGCCTGG TGGGCCTCCG CATCGTGTTC 2100
GCCGTGCTGA GCATCGTGAA CCGCGTGCGC CAGGGCTACA GCCCCCTGAG CCTCCAGACC 2160
CGGCCCCCCG TGCCGCGCGG GCCCGACCGC CCCGAGGGCA TCGAGGAGGA GGGCGGCGAG 2220
CGCGACCGCG ACACCAGCGG CAGGCTCGTG CACGGCTTCC TGGCGATCAT CTGGGTCGAC 2280
CTCCGCAGCC TGTTCCTGTT CAGCTACCAC CACCGCGACC TGCTGCTGAT CGCCGCCCGC 2340
ATCGTGGAAC TCCTAGGCCG CCGCGGCTGG GAGGTGCTGA AGTACTGGTG GAACCTCCTC 2400
CAGTATTGGA GCCAGGAGCT GAAGTCCAGC GCCGTGAGCC TGCTGAACGC CACCGCCATC 2460
GCCGTGGCCG AGGGCACCGA CCGCGTGATC GAGGTGCTCC AGAGGGCCGG GAGGGCGATC 2520
CTGCACATCC CCACCCGCAT CCGCCAGGGG CTCGAGAGGG CGCTGCTGTA A          2571
```

SEQ. I.D. NO. 11 - Complete Sequence of pH4DOZENEGS

```
CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA 60
CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT TCCTTTCTCG 120
CCACGTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG GCTCCCTTTA GGGTTCCGAT 180
TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG 240
GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA 300
GTGGACTCTT GTTCCAAACT GGAACAACAC TCAACCCTAT CTCGGTCTAT TCTTTTGATT 360
TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAAA TGAGCTGATT TAACAAAAAT 420
TTAACGCGAA TTTTAACAAA ATATTAACGC TTACAATTTC CATTCGCCAT TCAGGCTGCG 480
```

```
CAACTGTTGG GAAGGGCGAT CGGTGCGGGC CTCTTCGCTA TTACGCCAGC TGGCGAAAGG 540
GGGATGTGCT GCAAGGCGAT TAAGTTGGGT AACGCCAGGG TTTTCCCAGT CACGACGTTG 600
TAAAACGACG GCCAGTGAGC GCGCGTAATA CGACTCACTA TAGGGCGAAT TGGAGCTCCA 660
CCGCGGTGGC GGCCGCTCTA GAGTCCGTTA CATAACTTAC GGTAAATGGC CCGCCTGGCT 720
GACCGCCCAA CGACCCCGC CCATTGACGT CAATAATGAC GTATGTTCCC ATAGTAACGC 780
CAATAGGGAC TTTCCATTGA CGTCAATGGG TGGAGTATTT ACGGTAAACT GCCCACTTGG 840
CAGTACATCA AGTGTATCAT ATGCCAAGTA CGCCCCCTAT TGACGTCAAT GACGGTAAAT 900
GGCCCGCCTG GCATTATGCC CAGTACATGA CCTTATGGGA CTTTCCTACT TGGCAGTACA 960
TCTACGTATT AGTCATCGCT ATTACCATGG TGATGCGGTT TTGGCAGTAC ATCAATGGGC 1020
GTGGATAGCG GTTTGACTCA CGGGGATTTC CAAGTCTCCA CCCCATTGAC GTCAATGGGA 1080
GTTTGTTTTG GCACCAAAAT CAACGGGACT TTCCAAAATG TCGTAACAAC TCCGCCCCAT 1140
TGACGCAAAT GGGCGGTAGG CGTGTACGGT GGGAGGTCTA TATAAGCAGA GCTCGTTTAG 1200
TGAACCGGTC TCTCTGGTTA GACCAGATCT GAGCCTGGGA GCTCTCTGGC TAACTAGGGA 1260
ACCCACTGCT TAAGCCTCAA TAAAGCTTGC CTTGAGTGCT TCAAGTAGTG TGTGCCCGTC 1320
TGTTGTGTGA CTCTGGTAAC TAGAGATCCC TCAGACCCTT TTAGTCAGTG TGGAAAATCT 1380
CTAGCAGTGG CGCCCGAACA GGGACTTGAA AGCGAAAGGG AAACCAGAGG AGCTCTCTCG 1440
ACGCAGGACT CGGCTTGCTG AAGCGCGCAC GGCAAGAGGC GAGGGGCGGC GACTGGTGAG 1500
TACGCCAAAA ATTTTGACTA GCGGAGGCTA GAAGGAGAGA GATGGGTGCG AGAGCGTCAG 1560
TATTAAGCGG GGGAGAATTA GATCGCGATG GGAAAAAATT CGGTTAAGGC CAGGGGGAAA 1620
GAAAAAATAT AAATTAAAAC ATATAGTATG GGCAAGCAGG GAGCTAGAAC GATTCGCAGT 1680
TAATCCTGGC CTGTTAGAAA CATCAGAAGG CTGTAGACAA ATACTGGGAC AGCTACAACC 1740
ATCCCTTCAG ACAGGATCAG AAGAACTTAG ATCATTATAT AATACAGTAG CAACCCTCTA 1800
TTGTGTGCAT CAAAGGTTGA GATAAAAGAC ACCAAGGAAG CTTTAGACAA GATAGAGGGA 1860
GAGCAAAACA AAAGTAAGAA AAAAGCACAG CAAGCAGCAG CTGACACAGG ACACAGCAAT 1920
CAGGTCAGCC AAAATTACCC TATAGTGCAG AACATCCAGG GGCAAATGGT ACATCAGGCC 1980
ATATCACCTA GAACTTTAAA TGCATGGGTA AAAGTAGTAG AAGAGAAGGC TTTCAGCCCA 2040
GAAGTGATAC CCATGTTTTC AGCATTATCA GAAGGAGCCA CCCCACAAGA TTTAAACACC 2100
ATGCTAAACA CAGTGGGGGG ACATCAAGCA GCCATGCAAA TGTTAAAAGA GACCATCAAT 2160
GAGGAAGCTG CAGGAATTCG CCTAAAACTG CTTGTACCAA TTGCTATTGT AAAAAGTGTT 2220
GCTTTCATTG CCAAGTTTGT TTCATAACAA AAGCCTTAGG CATCTCCTAT GGCAGGAAGA 2280
AGCGGAGACA GCGACGAAGA GCTCATCAGA ACAGTCAGAC TCATCAAGCT TCTCTATCAA 2340
AGCAGTAAGT AGTACATGTA ACGCAACCTA TACCAATAGT AGCAATAGTA GCATTAGTAG 2400
TAGCAATAAT AATAGCAATA GTTGTGTGGT CCATAGTAAT CATAGAATAT AGGAAAATAT 2460
TAAGACAAAG AAAAATAGAC AGGTTAATTG ATAGACTAAT AGAAAGAGCA GAAGACAGTG 2520
GCAATGAGAG TGAAGGAGAA ATATCAGCAC TTGTGGAGAT GGGGGTGGAG ATGGGGCACC 2580
ATGCTCCTTG GGATGTTGAT GATCTGTAGT GCTACAGAAA AATTGTGGGT CACAGTCTAT 2640
TATGGGGTAC CTGTGTGGAA GGAAGCAACC ACCACTCTAT TTTGTGCATC AGATGCTAAA 2700
GCATAGATCT TCAGACTTGG AGGAGGAGAT ATGAGGGACA ATTGGAGAAG TGAATTATAT 2760
AAATATAAAG TAGTAAAAAT TGAACCATTA GGAGTAGCAC CCACCAAGGC AAAGAGAAGA 2820
GTGGTGCAGA GAGAAAAAAG AGCAGTGGGA ATAGGAGCTT TGTTCCTTGG GTTCTTGGGA 2880
GCAGCAGGAA GCACTATGGG CGCAGCGTCA ATGACGCTGA CGGTACAGGC CAGACAATTA 2940
TTGTCTGGTA TAGTGCAGCA GCAGAACAAT TTGCTGAGGG CTATTGAGGC GCAACAGCAT 3000
CTGTTGCAAC TCACAGTCTG GGGCATCAAG CAGCTCCAGG CAAGAATCCT GGCTGTGGAA 3060
AGATACCTAA AGGATCAACA GCTCCTGGGG ATTTGGGGTT GCTCTGGAAA ACTCATTTGC 3120
ACCACTGCTG TGCCTTGGAA TGCTAGTTGG AGTAATAAAT CTCTGGAACA GATCTGGAAT 3180
CACACGACCT GGATGGAGTG GGACAGAGAA ATTAACAATT ACACAAGCTT AATACACTCC 3240
TTAATTGAAG AATCGCAAAA CCAGCAAGAA AAGAATGAAC AAGAATTATT GGAATTAGAT 3300
AAATGGGCAA GTTTGTGGAA TTGGTTTAAC ATAACAAATT GGCTGTGGTA TATAAAATTA 3360
TTCATAATGA TAGTAGGAGG CTTGGTAGGT TTAAGAATAG TTTTTGCTGT ACTTTCTATA 3420
GTGAATAGAG TTAGGCAGGG ATATTCACCA TTATCGTTTC AGACCCACCT CCCAACCCCG 3480
AGGGGACCCG ACAGGCCCGA AGGAATAGAA GAAGAAGGTG GAGAGAGAGA CAGAGACAGA 3540
TCCATTCGAT TAGTGAACGG ATCCTTGGCA CTTATCTGGG ACGATCTGCG GAGCCTGTGC 3600
CTCTTCAGCT ACCACCGCTT GAGAGACTTA CTCTTGATTG TAACGAGGAT TGTGGAACTT 3660
CTGGGACGCA GGGGGTGGGA AGCCCTCAAA TATTGGTGGA ATCTCCTACA GTATTGGAGT 3720
CAGGAACTAA AGAATAGTGC TGTTAGCTTG CTCAATGCCA CAGCCATAGC AGTAGCTGAG 3780
GGGACAGATA GGGTTATAGA AGTAGTACAA GGAGCTTGTA GAGCTATTCG CCACATACCT 3840
AGAAGAATAA GACAGGGCTT GGAAAGGATT TTGCTATAAG ATGGGTGGCA AGTGGTCAAA 3900
AAGTAGTGTG ATTGGATGGC CTACTGTAAG GGAAAGAATG AGACGAGCTG AGCCAGCAGC 3960
AGATAGGGTG GGAGCAGCAT CTCGACGCTG CAGGAGTGGG GAGGCACGAT GGCCGCTTTG 4020
GTCGAGGCGG ATCCGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT 4080
TGGCTATTGG CCATTGCATA CGTTGTATCC ATATCATAAT ATGTACATTT ATATTGGCTC 4140
ATGTCCAACA TTACCGCCAT GTTGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT 4200
TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA 4260
TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT 4320
TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA 4380
```

```
AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT 4440
CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC 4500
TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA 4560
GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT 4620
TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA 4680
CAACTCCGCC CCATTGACGC AAATGGGCGG TAGGCATGTA CGGTGGGAGG TCTATATAAG 4740
CAGAGCTCGT TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT 4800
CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCCC AAGCTTCAGC TGCTCGAGCC 4860
CGGGGATGAC GTCATCGACT TCGAAGGTTC GAATCCTTCT ACTGCCACCA TTTTTTCTCT 4920
ACGTCATCGA CTTCGAAGGT TCGAATCCTT CCCTGTCCAC CAGTCGAGTA TTACGTCATC 4980
GACTTCGAAG GTTCGAATCC TTCTAGATTC ACCATTTTTT AGGAACGTCA TCGACTTCGA 5040
AGGTTCGAAT CCTTCCAGTT CCACCAGTCG AGGCCAACGT CATCGACTTC GAAGGTTCGA 5100
ATCCTTCTCT TCCCACCATT TTTTTTCCAC GTCATCGACT TCGAAGGTTC GAATCCTTCG 5160
GGGCCCACCA GTCGAGGGCT ACGTCATCGA CTTCGAAGGT TCGAATCCTT CTTGCTTCAC 5220
CATTTTTTCT GAACGTCATC GACTTCGAAG GTTCGAATCC TTCTGCTGTC ACCAGTCGAG 5280
TATAACGTCA TCGACTTCGA AGGTTCGAAT CCTTCACCGG TCACCATTTT TTTATAACGT 5340
CATCGACTTC GAAGGTTCGA ATCCTTCTTC TTACACCAGT CGAGGTACAC GTCATCGACT 5400
TCGAAGGTTC GAATCCTTCG TAGTTCACCA TTTTTTGTGC ACGTCATCGA CTTCGAAGGT 5460
TCGAATCCTT CTAGGCCCAC CAGTCGACGC ATGCCTGCAG GTCGAGGTCG ATACCGTCGA 5520
GACCTAGAAA AACATGGAGC AATCACAAGT AGCAATACAG CAGCTACCAA TGCTGATTGT 5580
GCCTGGCTAG AAGCACAAGA GGAGGAGGAG GTGGGTTTTC CAGTCACACC TCAGGTACCT 5640
TTAAGACCAA TGACTTACAA GGCAGCTGTA GATCTTAGCC ACTTTTTAAA AGAAAAGGGG 5700
GGACTGGAAG GGCTAATTCA CTCCCAACGA AGACAAGATA TCCTTGATCT GTGGATCTAC 5760
CACACACAAG GCTACTTCCC TGATTGGCAG AACTACACAC CAGGGCCAGG GATCAGATAT 5820
CCACTGACCT TTGGATGGTG CTACAAGCTA GTACCAGTTG AGCAAGAGAA GGTAGAAGAA 5880
GCCAATGAAG GAGAGAACAC CCGCTTGTTA CACCCTGTGA GCCTGCATGG GATGGATGAC 5940
CCGGAGAGAG AAGTATTAGA GTGGAGGTTT GACAGCCGCC TAGCATTTCA TCACATGGCC 6000
CGAGAGCTGC ATCCGGAGTA CTTCAAGAAC TGCTGACATC GAGCTTGCTA CAAGGGACTT 6060
TCCGCTGGGG ACTTTCCAGG GAGGCGTGGC CTGGGCGGGA CTGGGGAGTG GCGAGCCCTC 6120
AGATGCTGCA TATAAGCAGC TGCTTTTTGC CTGTACTGGG TCTCTCTGGT TAGACCAGAT 6180
CTGAGCCTGG GAGCTCTCTG GCTAACTAGG GAACCCACTG CTTAAGCCTC AATAAAGCTT 6240
GCCTTGAGTG CTTCAAGTAG TGTGTGCCCG TCTGTTGTGT GACTCTGGTA ACTAGAGATC 6300
CCTCAGACCC TTTTAGTCAG TGTGGAAAAT CTCTAGCAGT CGAGGGGGGG CCCGGTACCC 6360
AGCTTTTGTT CCCTTTAGTG AGGGTTAATT GCGCGCTTGG CGTAATCATG GTCATAGCTG 6420
TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA 6480
AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA 6540
CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC 6600
GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT CCTCGCTCAC TGACTCGCTG 6660
CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA 6720
TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC 6780
AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCTGACGAG 6840
CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC 6900
CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT GCCGCTTACC 6960
GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCATAG CTCACGCTGT 7020
AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC 7080
GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA 7140
CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA 7200
GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA 7260
TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA 7320
TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG 7380
CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG 7440
TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC 7500
TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT 7560
TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT 7620
CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GAGGGCTTA 7680
CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA 7740
TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC 7800
GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT 7860
AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT 7920
ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG 7980
TGCAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA 8040
GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA 8100
AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG 8160
CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT 8220
TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG 8280
```

```
CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT 8340
ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA 8400
ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC 8460
ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA 8520
CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC                        8560
```

SEQ. I.D. NO. 12 – pSYNGP2 – codon optimised HIV-1 gagpol with leader sequence

```
   1 GGGTCTCTCT GGTTAGACCA GATCTGAGCC TGGGAGCTCT CTGGCTAACT AGGGAACCCA
  61 CTGCTTAAGC CTCAATAAAG CTTGCCTTGA GTGCTTCAAG TAGTGTGTGC CCGTCTGTTG
 121 TGTGACTCTG GTAACTAGAG ATCCCTCAGA CCCTTTTAGT CAGTGTGGAA AATCTCTAGC
 181 AGTGGCGCCC GAACAGGGAC CTGAAAGCGA AAGGGAAACC AGAGCTCTCT CGACGCAGGA
 241 CTCGGCTTGC TGAAGCGCCC GCACGGCAAG AGGCGAGGGG CGGCGACTGG TGAGTACGCC
 301 AAAAATTTTG ACTAGCGGAG GCTAGAAGGA GAGAGATGGG CGCCCGCGCC AGCGTGCTGT
 361 CGGGCGGCGA GCTGGACCGC TGGGAGAAGA TCCGCCTGCG CCCCGGCGGC AAAAAGAAGT
 421 ACAAGCTGAA GCACATCGTG TGGGCCAGCC GCGAACTGGA GCGCTTCGCC GTGAACCCCG
 481 GGCTCCTGGA GACCAGCGAG GGGTGCCGCC AGATCCTCGG CCAACTGCAG CCCAGCCTGC
 541 AAACCGGCAG CGAGGAGCTG CGCAGCCTGT ACAACACCGT GGCCACGCTG TACTGCGTCC
 601 ACCAGCGCAT CGAAATCAAG GATACGAAAG AGGCCCTGGA TAAAATCGAA GAGGAACAGA
 661 ATAAGAGCAA AAAGAAGGCC CAACAGGCCG CAGCAGCAGC CGGCACAGGA AACAGCAGCA
 721 GCCAGAACTA CCCCATCGTG CAGAACATCC AGGGGCAGAT GGTGCACCAG GCCATCTCCC
 781 CCCGCACGCT GAACGCCTGG GTGAAGGTGG TGGAAGAGAA GGCTTTTAGC CCGGAGGTGA
 841 TACCCATGTT CTCAGCCCTG TCAGAGGGAG CCACCCCCCA AGATCTGAAC ACCATGCTCA
 901 ACACAGTGGG GGGACACCAG GCCGCCATGC AGATGCTGAA GGAGACCATC AATGAGGAGG
 961 CTGCCGAATG GGATCGTGTG CATCCGGTGC ACGCAGGGCC CATCGCACCG GGCCAGATGC
1021 GTGAGCCACG GGGCTCAGAC ATCGCCGGAA CGACTAGTAC CCTTCAGGAA CAGATCGGCT
1081 GGATGACCAA CAACCCACCC ATCCCGGTGG GAGAAATCTA CAAACGCTGG ATCATCCTGG
1141 GCCTGAACAA GATCGTGCGC ATGTATAGCC CTACCAGCAT CCTGGACATC CGCCAAGGCC
1201 CGAAGGAACC CTTTCGCGAC TACGTGGACC GGTTCTACAA AACGCTCCGC GCCGAGCAGG
1261 CTAGCCAGGA GGTGAAGAAC TGGATGACCG AAACCCTGCT GGTCCAGAAC GCGAACCCGG
1321 ACTGCAAGAC GATCCTGAAG GCCCTGGGCC CAGCGGCTAC CCTAGAGGAA ATGATGACCG
1381 CCTGTCAGGG AGTGGGCGGA CCCGGCCACA AGGCACGCGT CCTGGCTGAG GCCATGAGCC
1441 AGGTGACCAA CTCCGCTACC ATCATGATGC AGCGCGGCAA CTTTCGGAAC CAACGCAAGA
1501 TCGTCAAGTG CTTCAACTGT GGCAAAGAAG GCCACACAGC CCGCAACTGC AGGGCCCCTA
1561 GGAAAAAGGG CTGTTGGAAA TGTGGAAAGG AAGGACACCA AATGAAAGAT TGTACTGAGA
1621 GACAGGCTAA TTTTTTAGGG AAGATCTGGC CTTCCCACAA GGGAAGGCCA GGGAATTTTC
1681 TTCAGAGCAG ACCAGAGCCA ACAGCCCCAC CAGAAGAGAG CTTCAGGTTT GGGGAAGAGA
1741 CAACAACTCC CTCTCAGAAG CAGGAGCCGA TAGACAAGGA ACTGTATCCT TTAGCTTCCC
1801 TCAGATCACT CTTTGGCAGC GACCCCTCGT CACAATAAAG ATAGGGGGGC AGCTCAAGGA
1861 GGCTCTCCTG GACACCGGAG CAGACGACAC CGTGCTGGAG GAGATGTCGT TGCCAGGCGA
1921 CTGGAAGCCG AAGATGATCG GGGGAATCGG CGGTTTCATC AAGGTGCGCC AGTATGACCA
1981 GATCCTCATC GAAATCTGCG GCCACAAGGC TATCGGTACC GTGCTGGTGG GCCCCACACC
2041 CGTCAACATC ATCGGACGCA ACCTGTTGAC GCAGATCGGT TGCACGCTGA ACTTCCCCAT
2101 TAGCCCTATC GAGACGGTAC CGGTGAAGCT GAAGCCCGGG ATGGACGGCC CGAAGGTCAA
2161 GCAATGGCCA TTGACAGAGG AGAAGATCAA GGCACTGGTG GAGATTTGCA CAGAGATGGA
2221 AAAGGAAGGG AAAATCTCCA AGATTGGGCC TGAGAACCCG TACAACACGC CGGTGTTCGC
2281 AATCAAGAAG AAGGACTCGA CGAAATGGCG CAAGCTGGTG GACTTCCGCG AGCTGAACAA
2341 GCGCACGCAA GACTTCTGGG AGGTTCAGCT GGGCATCCCG CACCCCGCAG GCTGAAGAA
2401 GAAGAAATCC GTGACCGTAC TGGATGTGGG TGATGCCTAC TTCTCCGTTC CCTGGACGA
2461 AGACTTCAGG AAGTACACTG CCTTCACAAT CCCTTCGATC AACAACGAGA CACCGGGGAT
2521 TCGATATCAG TACAACGTGC TGCCCCAGGG CTGGAAAGGC TCTCCCGCAA TCTTCCAGAG
2581 TAGCATGACC AAAATCCTGG AGCCTTTCCG CAAACAGAAC CCCGACATCG TCATCTATCA
2641 GTACATGGAT GACTTGTACG TGGGCTCTGA TCTAGAGATA GGGCAGCACC GCACCAAGAT
2701 CGAGGAGCTG CGCCAGCACC TGTTGAGGTG GGGACTGACC ACACCCGACA AGAAGCACCA
2761 GAAGGAGCCT CCCTTCCTCT GGATGGGTTA CGAGCTGCAC CCTGACAAAT GGACCGTGCA
2821 GCCTATCGTG CTGCCAGAAG AAGACAGCTG GACTGTCAAC GACATACAGA AGCTGGTGGG
2881 GAAGTTGAAC TGGGCCAGTC AGATTTACCC AGGGATTAAG GTGAGGCAGC TGTGCAAACT
2941 CCTCCGCGGA ACCAAGGCAC TCACAGAGGT GATCCCCCTA ACCGAGGAGG CCGAGCTCGA
3001 ACTGGCAGAA AACCGAGAGA TCCTAAAGGA GCCCGTGCAC GGCGTGTACT ATGACCCCTC
3061 CAAGGACCTG ATCGCCGAGA TCCAGAAGCA GGGGCAAGGC CAGTGGACCT ATCAGATTTA
3121 CCAGGAGCCC TTCAAGAACC TGAAGACCGG CAAGTACGCC CGGATGAGGG GTGCCCACAC
3181 TAACGACGTC AAGCAGCTGA CCGAGGCCGT GCAGAAGATC ACCACCGAAA GCATCGTGAT
3241 CTGGGGAAAG ACTCCTAAGT TCAAGCTGCC CATCCAGAAG GAAACCTGGG AAACCTGGTG
3301 GACAGAGTAT TGGCAGGCCA CCTGGATTCC TGAGTGGGAG TTCGTCAACA CCCCTCCCCT
3361 GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATAGTG GGCGCCGAAA CCTTCTACGT
```

```
3421 GGATGGGGCC GCTAACAGGG AGACTAAGCT GGGCAAAGCC GGATACGTCA CTAACCGGGG
3481 CAGACAGAAG GTTGTCACCC TCACTGACAC CACCAACCAG AAGACTGAGC TGCAGGCCAT
3541 TTACCTCGCT TTGCAGGACT CGGGCCTGGA GGTGAACATC GTGACAGACT CTCAGTATGC
3601 CCTGGGCATC ATTCAAGCCC AGCCAGACCA GAGTGAGTCC GAGCTGGTCA ATCAGATCAT
3661 CGAGCAGCTG ATCAAGAAGG AAAAGGTCTA TCTGGCCTGG GTACCCGCCC ACAAAGGCAT
3721 TGGCGGCAAT GAGCAGGTCG ACAAGCTGGT CTCGGCTGGC ATCAGGAAGG TGCTATTCCT
3781 GGATGGCATC GACAAGGCCC AGGACGAGCA CGAGAAATAC CACAGCAACT GGCGGGCCAT
3841 GGCTAGCGAC TTCAACCTGC CCCCTGTGGT GGCCAAAGAG ATCGTGGCCA GCTGTGACAA
3901 GTGTCAGCTC AAGGGCGAAG CCATGCATGG CCAGGTGGAC TGTAGCCCCG GCATCTGGCA
3961 ACTCGATTGC ACCCATCTGG AGGGCAAGGT TATCCTGGTA GCCGTCCATG TGGCCAGTGG
4021 CTACATCGAG GCCGAGGTCA TTCCCGCCGA AACAGGGCAG GAGACAGCCT ACTTCCTCCT
4081 GAAGCTGGCA GGCCGGTGGC CAGTGAAGAC CATCCATACT GACAATGGCA GCAATTTCAC
4141 CAGTGCTACG GTTAAGGCCG CCTGCTGGTG GGCGGGAATC AAGCAGGAGT TCGGGATCCC
4201 CTACAATCCC CAGAGTCAGG GCGTCGTCGA GTCTATGAAT AAGGAGTTAA AGAAGATTAT
4261 CGGCCAGGTC AGAGATCAGG CTGAGCATCT CAAGACCGCG GTCCAAATGG CGGTATTCAT
4321 CCACAATTTC AAGCGGAAGG GGGGGATTGG GGGGTACAGT GCGGGGGAGC GGATCGTGGA
4381 CATCATCGCG ACCGACATCC AGACTAAGGA GCTGCAAAAG CAGATTACCA AGATTCAGAA
4441 TTTCCGGGTC TACTACAGGG ACAGCAGAAA TCCCCTCTGG AAAGGCCCAG CGAAGCTCCT
4501 CTGGAAGGGT GAGGGGGCAG TAGTGATCCA GGATAATAGC GACATCAAGG TGGTGCCCAG
4561 AAGAAAGGCG AAGATCATTA GGGATTATGG CAAACAGATG GCGGGTGATG ATTGCGTGGC
4621 GAGCAGACAG GATGAGGATT AG
```

SEQ. I.D. NO. 13 – pSYNGP3 – codon optimised HIV-1 gagpol with leader sequence from the major splice donor

```
   1 GTGAGTACGC CAAAAATTTT GACTAGCGGA GGCTAGAAGG AGAGAGATGG GCGCCCGCGC
  61 CAGCGTGCTG TCGGGCGGCG AGCTGGACCG CTGGGAGAAG ATCCGCCTGC GCCCCGGCGG
 121 CAAAAAGAAG TACAAGCTGA AGCACATCGT GTGGGCCAGC CGCGAACTGG AGCGCTTCGC
 181 CGTGAACCCC GGGCTCCTGG AGACCAGCGA GGGGTGCCGC CAGATCCTCG GCCAACTGCA
 241 GCCCAGCCTG CAAACCGGCA GCGAGGAGCT GCGCAGCCTG TACAACACCG TGGCCACGCT
 301 GTACTGCGTC CACCAGCGCA TCGAAATCAA GGATACGAAA GAGGCCCTGG ATAAAATCGA
 361 AGAGGAACAG AATAAGAGCA AAAGAAGGC CCAACAGGCC GCCGCGGACA CCGGACACAG
 421 CAACCAGGTC AGCCAGAACT ACCCCATCGT GCAGAACATC CAGGGGCAGA TGGTGCACCA
 481 GGCCATCTCC CCCCGCACGC TGAACGCCTG GGTGAAGGTG GTGGAAGAGA GGCTTTTAG
 541 CCCGGAGGTG ATACCCATGT TCTCAGCCCT GTCAGAGGGA GCCACCCCCC AAGATCTGAA
 601 CACCATGCTC AACACAGTGG GGGACACCA GGCCGCCATG CAGATGCTGA AGGAGACCAT
 661 CAATGAGGAG GCTGCCGAAT GGGATCGTGT GCATCCGGTG CACGCAGGGC CCATCGCACC
 721 GGGCCAGATG CGTGAGCCAC GGGGCTCAGA CATCGCCGGA ACGACTAGTA CCCTTCAGGA
 781 ACAGATCGGC TGGATGACCA ACAACCCACC CATCCCGGTG GGAGAAATCT ACAAACGCTG
 841 GATCATCCTG GGCCTGAACA AGATCGTGCG CATGTATAGC CCTACCAGCA TCCTGGACAT
 901 CCGCCAAGGC CCGAAGGAAC CCTTTCGCGA CTACGTGGAC CGGTTCTACA AAACGCTCCG
 961 CGCCGAGCAG GCTAGCCAGG AGGTGAAGAA CTGGATGACC GAAACCCTGC TGGTCCAGAA
1021 CGCGAACCCG GACTGCAAGA CGATCCTGAA GGCCCTGGGC CCAGCGGCTA CCCTAGAGGA
1081 AATGATGACC GCCTGTCAGG GAGTGGGCGG ACCCGGCCAC AAGGCACGCG TCCTGGCTGA
1141 GGCCATGAGC CAGGTGACCA ACTCCGCTAC CATCATGATG CAGCGCGGCA ACTTTCGGAA
1201 CCAACGCAAG ATCGTCAAGT GCTTCAACTG TGGCAAAGAA GGGCACACAG CCCGCAACTG
1261 CAGGGCCCCT AGGAAAAAGG CTGTTGGAA ATGTGGAAAG GAAGGACACC AAATGAAAGA
1321 TTGTACTGAG AGACAGGCTA ATTTTTTAGG GAAGATCTGG CCTTCCCACA GGGGAAGGCC
1381 AGGGAATTTT CTTCAGAGCA GACCAGAGCC AACAGCCCCA CCAGAAGAGA GCTTCAGGTT
1441 TGGGGAAGAG ACAACAACTC CCTCTCAGAA GCAGGAGCCG ATAGACAAG AACTGTATCC
1501 TTTAGCTTCC CTCAGATCAC TCTTTGGCAG CGACCCCTCG TCACAATAAA GATAGGGGGG
1561 CAGCTCAAGG AGGCTCTCCT GGACACCGGA GCAGACGACA CCGTGCTGGA GGAGATGTCG
1621 TTGCCAGGCC GCTGGAAGCC GAAGATGATC GGGGGAATCG GCGGTTTCAT CAAGGTGCGC
1681 CAGTATGACC AGATCCTCAT CGAAATCTGC GGCCACAAGG CTATCGGTAC CGTGCTGGTG
1741 GGCCCCACAC CCGTCAACAT CATCGGACGC AACCTGTTGA CGCAGATCGG TTGCACGCTG
1801 AACTTCCCCA TTAGCCCTAT CGAGACGGTA CCGGTGAAGC TGAAGCCCGG GATGGACGGC
1861 CCGAAGGTCA AGCAATGGCC ATTGACAGAG GAGAAGATCA AGGCACTGGT GGAGATTTGC
1921 ACAGAGATGG AAAAGGAAGG GAAAATCTCC AAGATTGGGC CTGAGAACCC GTACAACACG
1981 CCGGTGTTCG CAATCAAGAA GAAGGACTCG ACGAAATGGC GCAAGCTGGT GGACTTCCGC
2041 GAGCTGAACA AGCGCACGCA AGACTTCTGG GAGGTTCAGC TGGGCATCCC GCACCCCGCA
2101 GGGCTGAAGA AGAAGAAATC CGTGACCGTA CTGGATGTGG GTGATGCCTA CTTCTCCGTT
2161 CCCCTGGACG AAGACTTCAG GAAGTACACT GCCTTCACAA TCCCTTCGAT CAACAACGAG
2221 ACACCGGGGA TTCGATATCA GTACAACGTG CTGCCCCAGG GCTGGAAAGG CTCTCCCGCA
```

```
2281 ATCTTCCAGA GTAGCATGAC CAAAATCCTG GAGCCTTTCC GCAAACAGAA CCCCGACATC
2341 GTCATCTATC AGTACATGGA TGACTTGTAC GTGGGCTCTG ATCTAGAGAT AGGGCAGCAC
2401 CGCACCAAGA TCGAGGAGCT GCGCCAGCAC CTGTTGAGGT GGGGACTGAC CACACCCGAC
2461 AAGAAGCACC AGAAGGAGCC TCCCTTCCTC TGGATGGGTT ACGAGCTGCA CCCTGACAAA
2521 TGGACCGTGC AGCCTATCGT GCTGCCAGAG AAAGACAGCT GGACTGTCAA CGACATACAG
2581 AAGCTGGTGG GGAAGTTGAA CTGGGCCAGT CAGATTTACC CAGGGATTAA GGTGAGGCAG
2641 CTGTGCAAAC TCCTCCGCGG AACCAAGGCA CTCACAGAGG TGATCCCCCT AACCGAGGAG
2701 GCCGAGCTCG AACTGGCAGA AAACCGAGAG ATCCTAAAGG AGCCCGTGCA CGGCGTGTAC
2761 TATGACCCCT CCAAGGACCT GATCGCCGAG ATCCAGAAGC AGGGGCAAGG CCAGTGGACC
2821 TATCAGATTT ACCAGGAGCC CTTCAAGAAC CTGAAGACCG GCAAGTACGC CCGGATGAGG
2881 GGTGCCCACA CTAACGACGT CAAGCAGCTG ACCGAGGCCG TGCAGAAGAT CACCACCGAA
2941 AGCATCGTGA TCTGGGGAAA GACTCCTAAG TTCAAGCTGC CCATCCAGAA GGAAACCTGG
3001 GAAACCTGGT GGACAGAGTA TTGGCAGGCC ACCTGGATTC CTGAGTGGGA GTTCGTCAAC
3061 ACCCCTCCCC TGGTGAAGCT GTGGTACCAG CTGGAGAAGG AGCCCATAGT GGGCGCCGAA
3121 ACCTTCTACG TGGATGGGGC CGCTAACAGG GAGACTAAGC TGGGCAAAGC CGGATACGTC
3181 ACTAACCGGG GCAGACAGAA GGTTGTCACC CTCACTGACA CCACCAACCA GAAGACTGAG
3241 CTGCAGGCCA TTTACCTCGC TTTGCAGGAC TCGGGCCTGG AGGTGAACAT CGTGACAGAC
3301 TCTCAGTATG CCCTGGGCAT CATTCAAGCC CAGCCAGACC AGAGTGAGTC CGAGCTGGTC
3361 AATCAGATCA TCGAGCAGCT GATCAAGAAG GAAAAGGTCT ATCTGGCCTG GGTACCCGCC
3421 CACAAAGGCA TTGGCGGCAA TGAGCAGGTC GACAAGCTGG TCTCGGCTGG CATCAGGAAG
3481 GTGCTATTCC TGGATGGCAT CGACAAGGCC CAGGACGAGC ACGAGAAATA CCACAGCAAC
3541 TGGCGGGCCA TGGCTAGCGA CTTCAACCTG CCCCCTGTGG TGGCCAAAGA GATCGTGGCC
3601 AGCTGTGACA AGTGTCAGCT CAAGGGCGAA GCCATGCATG GCCAGGTGGA CTGTAGCCCC
3661 GGCATCTGGC AACTCGATTG CACCCATCTG GAGGGCAAGG TTATCCTGGT AGCCGTCCAT
3721 GTGGCCAGTG GCTACATCGA GGCCGAGGTC ATTCCCGCCG AAACAGGGCA GGAGACAGCC
3781 TACTTCCTCC TGAAGCTGGC AGGCCGGTGG CCAGTGAAGA CCATCCATAC TGACAATGGC
3841 AGCAATTTCA CCAGTGCTAC GGTTAAGGCC GCCTGCTGGT GGGCGGGAAT CAAGCAGGAG
3901 TTCGGGATCC CCTACAATCC CCAGAGTCAG GGCGTCGTCG AGTCTATGAA TAAGGAGTTA
3961 AAGAAGATTA TCGGCCAGGT CAGAGATCAG GCTGAGCACC TCAAGACCGC GGTCCAAATG
4021 GCGGTATTCA TCCACAATTT CAAGCGGAAG GGGGGGATTG GGGGGTACAG TGCGGGGGAG
4081 CGGATCGTGG ACATCATCGC GACCGACATC CAGACTAAGG AGCTGCAAAA GCAGATTACC
4141 AAGATTCAGA ATTTCCGGGT CTACTACAGG GACAGCAGAA ATCCCCTCTG GAAAGGCCCA
4201 GCGAAGCTCC TCTGGAAGGG TGAGGGGGCA GTAGTGATCC AGGATAATAG CGACATCAAG
4261 GTGGTGCCCA GAAGAAAGGC GAAGATCATT AGGGATTATG GCAAACAGAT GGCGGGTGAT
4321 GATTGCGTGG CGAGCAGACA GGATGAGGAT TAG
```

SEQ. I.D. NO. 14 – pSYNGP4 – codon optimised HIV-1 gagpol with 20 bp of the leader sequence of HIV-1, upstream of the start codon of ATG.

```
   1 CGGAGGCTAG AAGGAGAGAG ATGGGCGCCC GCGCCAGCGT GCTGTCGGGC GGCGAGCTGG
  61 ACCGCTGGGA GAAGATCCGC CTGCGCCCCG GCGGCAAAAA GAAGTACAAG CTGAAGCACA
 121 TCGTGTGGGC CAGCCGCGAA CTGGAGCGCT TCGCCGTGAA CCCCGGGCTC CTGGAGACCA
 181 GCGAGGGGTG CCGCCAGATC CTCGGCCAAC TGCAGCCCAG CCTGCAAACC GGCAGCGAGG
 241 AGCTGCGCAG CCTGTACAAC ACCGTGGCCA CGCTGTACTG CGTCCACCAG CGCATCGAAA
 301 TCAAGGATAC GAAAGAGGCG CTGGATAAAA TCGAAGAGGA ACAGAATAAG AGCAAAAAGA
 361 AGGCCCAACA GGCCGCGCG GACACCGGAC ACAGCAACCA GGTCAGCCAG AACTACCCCA
 421 TCGTGCAGAA CATCCAGGGG CAGATGGTGC ACCAGGCCAT CTCCCCCCGC ACGCTGAACG
 481 CCTGGGTGAA GGTGGTGGAA GAGAAGGCTT TTAGCCCGGA GGTGATACCC ATGTTCTCAG
 541 CCCTGTCAGA GGGAGCCACC CCCCAAGATC TGAACACCAT GCTCAACACA GTGGGGGGGAC
 601 ACCAGGCCGC CATGCAGATG CTGAAGGAGA CCATCAATGA GGAGGCTGCC GAATGGGATC
 661 GTGTGCATCC GGTGCACGCA GGGCCCATCG CACCGGGCCA GATGCGTGAG CCACGGGGCT
 721 CAGACATCGC CGGAACGACT AGTACCCTTC AGGAACAGAT CGGCTGGATG ACCAACAACC
 781 CACCCATCCC GGTGGGAGAA ATCTACAAAC GCTGGATCAT CCTGGGCCTG AACAAGATCG
 841 TGCGCATGTA TAGCCCTACC AGCATCCTGG ACATCCGCCA AGGCCCGAAG GAACCCTTTC
 901 GCGACTACGT GGACCGGTTC TACAAAACGC TCCGCGCCGA GCAGGCTAGC CAGGAGGTGA
 961 AGAACTGGAT GACCGAAACC CTGCTGGTCC AGAACGCGAA CCCGGACTGC AAGACGATCC
1021 TGAAGGCCCT GGGCCCAGCG GCTACCCTAG AGGAAATGAT GACCGCCTGT CAGGGAGTGG
1081 GCGGACCCGG CCACAAGGCA CGCGTCCTGG CTGAGGCCAT GAGCCAGGTG ACCAACTCCG
1141 CTACCATCAT GATGCAGCGC GGCAACTTTC GGAACCAACG CAAGATCGTC AAGTGCTTCA
1201 ACTGTGGCAA AGAAGGGCAC ACAGCCCGCA ACTGCAGGGC CCCTAGGAAA AAGGGCTGTT
1261 GGAAATGTGG AAAGGAAGGA CACCAAATGA AAGATTGTAC TGAGAGACAG GCTAATTTTT
1321 TAGGGAAGAT CTGGCCTTCC CACAAGGGAA GGCCAGGGAA TTTTCTTCAG AGCAGACCAG
1381 AGCCAACAGC CCCACCAGAA GAGAGCTTCA GGTTTGGGGA AGAGACAACA ACTCCCTCTC
1441 AGAAGCAGGA GCCGATAGAC AAGGAACTGT ATCCTTTAGC TTCCCTCAGA TCACTCTTTG
```

```
1501 GCAGCGACCC CTCGTCACAA TAAAGATAGG GGGGCAGCTC AAGGAGGCTC TCCTGGACAC
1561 CGGAGCAGAC GACACCGTGC TGGAGGAGAT GTCGTTGCCA GGCCGCTGGA AGCCGAAGAT
1621 GATCGGGGGA ATCGGCGGTT TCATCAAGGT GCGCCAGTAT GACCAGATCC TCATCGAAAT
1681 CTGCGGCCAC AAGGCTATCG GTACCGTGCT GGTGGGCCCC ACACCCGTCA ACATCATCGG
1741 ACGCAACCTG TTGACGCAGA TCGGTTGCAC GCTGAACTTC CCCATTAGCC CTATCGAGAC
1801 GGTACCGGTG AAGCTGAAGC CCGGGATGGA CGGCCCGAAG GTCAAGCAAT GGCCATTGAC
1861 AGAGGAGAAG ATCAAGGCAC TGGTGGAGAT TTGCACAGAG ATGGAAAAGG AAGGGAAAAT
1921 CTCCAAGATT GGGCCTGAGA ACCCGTACAA CACGCCGGTG TTCGCAATCA AGAAGAAGGA
1981 CTCGACGAAA TGGCGCAAGC TGGTGGACTT CCGCGAGCTG AACAAGCGCA CGCAAGACTT
2041 CTGGGAGGTT CAGCTGGGCA TCCCGCACCC CGCAGGGCTG AAGAAGAAGA AATCCGTGAC
2101 CGTACTGGAT GTGGGTGATG CCTACTTCTC CGTTCCCCTG GACGAAGACT TCAGGAAGTA
2161 CACTGCCTTC ACAATCCCTT CGATCAACAA CGAGACACCG GGGATTCGAT ATCAGTACAA
2221 CGTGCTGCCC CAGGGCTGGA AAGGCTCTCC CGCAATCTTC CAGAGTAGCA TGACCAAAAT
2281 CCTGGAGCCT TTCCGCAAAC AGAACCCCGA CATCGTCATC TATCAGTACA TGGATGACTT
2341 GTACGTGGGC TCTGATCTAG AGATAGGGCA GCACCGCACC AAGATCGAGG AGCTGCGCCA
2401 GCACCTGTTG AGGTGGGGAC TGACCACACC CGACAAGAAG CACCAGAAGG AGCCTCCCTT
2461 CCTCTGGATG GGTTACGAGC TGCACCCTGA CAAATGGACC GTGCAGCCTA TCGTGCTGCC
2521 AGAGAAAGAC AGCTGGACTG TCAACGACAT ACAGAAGCTG GTGGGGAAGT TGAACTGGGC
2581 CAGTCAGATT TACCCAGGGA TTAAGGTGAG GCAGCTGTGC AAACTCCTCC GCGGAACCAA
2641 GGCACTCACA GAGGTGATCC CCCTAACCGA GGAGGCCGAG CTCGAACTGG CAGAAAACCG
2701 AGAGATCCTA AAGGAGCCCG TGCACGGCGT GTACTATGAC CCCTCCAAGG ACCTGATCGC
2761 CGAGATCCAG AAGCAGGGGC AAGGCCAGTG GACCTATCAG ATTTACCAGG AGCCCTTCAA
2821 GAACCTGAAG ACCGGCAAGT ACGCCGGGAT GAGGGGTGCC CACACTAACG ACGTCAAGCA
2881 GCTGACCGAG GCCGTGCAGA AGATCACCAC CGAAAGCATC GTGATCTGGG GAAAGACTCC
2941 TAAGTTCAAG CTGCCCATCC AGAAGGAAAC CTGGGAAACC TGGTGGACAG AGTATTGGCA
3001 GGCCACCTGG ATTCCTGAGT GGGAGTTCGT CAACACCCCT CCCCTGGTGA AGCTGTGGTA
3061 CCAGCTGGAG AAGGAGCCCA TAGTGGGCGC CGAAACCTTC TACGTGGATG GGGCCGCTAA
3121 CAGGGAGACT AAGCTGGGCA AAGCCGGATA CGTCACTAAC CGGGGCAGAC AGAAGGTTGT
3181 CACCCTCACT GACACCACCA ACCAGAAGAC TGAGCTGCAG GCCATTTACC TCGCTTTGCA
3241 GGACTCGGGC CTGGAGGTGA ACATCGTGAC AGACTCTCAG TATGCCCTGG GCATCATTCA
3301 AGCCCAGCCA GACCAGAGTG AGTCCGAGCT GGTCAATCAG ATCATCGAGC AGCTGATCAA
3361 GAAGGAAAAG GTCTATCTGG CCTGGGTACC CGCCCACAAA GGCATTGGCG GCAATGAGCA
3421 GGTCGACAAG CTGGTCTCGG CTGGCATCAG GAAGGTGCTA TTCCTGGATG GCATCGACAA
3481 GGCCCAGGAC GAGCACGAGA AATACCACAG CAACTGGCGG GCCATGGCTA GCGACTTCAA
3541 CCTGCCCCCT GTGGTGGCCA AAGAGATCGT GGCCAGCTGT GACAAGTGTC AGCTCAAGGG
3601 CGAAGCCATG CATGGCCAGG TGGACTGTAG CCCCGGCATC TGGCAACTCG ATTGCACCCA
3661 TCTGGAGGGC AAGGTTATCC TGGTAGCCGT CCATGTGGCC AGTGGCTACA TCGAGGCCGA
3721 GGTCATTCCC GCCGAAACAG GGCAGGAGAC AGCCTACTTC CTCCTGAAGC TGGCAGGCCG
3781 GTGGCCAGTG AAGACCATCC ATACTGACAA TGGCAGCAAT TTCACCAGTG CTACGGTTAA
3841 GGCCGCCTGC TGGTGGGCGG GAATCAAGCA GGAGTTCGGG ATCCCCTACA ATCCCCAGAG
3901 TCAGGGCGTC GTCGAGTCTA TGAATAAGGA GTTAAAGAAG ATTATCGGCC AGGTCAGAGA
3961 TCAGGCTGAG CATCTCAAGA CCGCGGTCCA AATGGCGGTA TTCATCCACA ATTTCAAGCG
4021 GAAGGGGGGG ATTGGGGGGT ACAGTGCGGG GGAGCGGATC GTGGACATCA TCGCGACCGA
4081 CATCCAGACT AAGGAGCTGC AAAAGCAGAT TACCAAGATT CAGAATTTCC GGGTCTACTA
4141 CAGGGACAGC AGAAATCCCC TCTGGAAAGG CCCAGCGAAG CTCCTCTGGA AGGGTGAGGG
4201 GGCAGTAGTG ATCCAGGATA TAGCGACAT CAAGGTGGTG CCCAGAAGAA AGGCGAAGAT
4261 CATTAGGGAT TATGGCAAAC AGATGGCGGG TGATGATTGC GTGGCGAGCA GACAGGATGA
4321 GGATTAG
```